

Crystal structure of the dimeric protein core of decorin, the archetypal small leucine-rich repeat proteoglycan

Paul G. Scott*, Paul A. McEwan[†], Carole M. Dodd*, Ernst M. Bergmann*[‡], Paul N. Bishop[†], and Jordi Bella^{†§}

*Department of Biochemistry and [†]Alberta Synchrotron Institute, University of Alberta, Edmonton, AB, Canada T6G 2H7; and [‡]Wellcome Trust Centre for Cell-Matrix Research, University of Manchester, Manchester M13 9PT, United Kingdom

Edited by Timothy A. Springer, Harvard Medical School, Boston, MA, and approved July 27, 2004 (received for review April 28, 2004)

Decorin is a ubiquitous extracellular matrix proteoglycan with a variety of important biological functions that are mediated by its interactions with extracellular matrix proteins, cytokines, and cell surface receptors. Decorin is the prototype of the family of small leucine-rich repeat proteoglycans and proteins (SLRPs), characterized by a protein core composed of leucine-rich repeats (LRRs), flanked by two cysteine-rich regions. We report here the crystal structure of the dimeric protein core of decorin, the best characterized member of the SLRP family. Each monomer adopts the curved solenoid fold characteristic of LRR domains, with a parallel β -sheet on the inside interwoven with loops containing short segments of β -strands, 3_{10} helices, and polyproline II helices on the outside. Two main features are unique to this structure. First, decorin dimerizes through the concave surfaces of the LRR domains, which have been implicated previously in protein-ligand interactions. The amount of surface buried in this dimer rivals the buried surfaces of some of the highest-affinity macromolecular complexes reported to date. Second, the C-terminal region adopts an unusual capping motif that involves a laterally extended LRR and a disulfide bond. This motif seems to be unique to SLRPs and has not been observed in any other LRR protein structure to date. Possible implications of these features for decorin ligand binding and SLRP function are discussed.

Decorin is a small extracellular matrix proteoglycan present in a variety of connective tissues, typically in association with or “decorating” collagen fibrils (1–3). It is involved in several fundamental biological functions, including the formation and/or organization of collagen fibrils (4, 5) and the modulation of cell adhesion mediated by fibronectin and thrombospondin (6). Decorin also modulates the activity of growth factors, such as transforming growth factor- β (7), and has other, transforming growth factor- β -independent effects on cell proliferation and behavior (8, 9).

Mammalian decorin contains a protein core and a single chondroitin/dermatan sulfate glycosaminoglycan (GAG) chain, attached to a serine residue near the N terminus (10). Decorin is the best characterized member of the growing family of small leucine-rich repeat proteoglycans and proteins (SLRPs) (3, 11), all having a domain of tandem leucine-rich repeats (LRRs), flanked on either side by clusters of conserved Cys residues. Most SLRPs have been grouped into three different classes on the basis of gene organization, amino acid sequence similarity, number of LRRs, and the spacing of Cys residues in the N-terminal segment. Thus, class I includes decorin, biglycan, and asporin; class II includes fibromodulin, osteoadherin, lumican, proline arginine-rich end LRR protein (PRELP), and keratocan; and class III includes opticin, osteoglycin/mimican, and epiphykan/PGLb (3, 12). Three further proteins, extracellular matrix 2 (ECM2), chondroadherin, and nctalopin, have LRR domains with significant homology to the SLRP family (12). The structural and functional similarities between different SLRPs suggest that they share biological functions. For instance, several SLRPs are known to regulate collagen fibrillogenesis, and there is

evidence that they are able to compensate for each other in studies on knockout mice (11). Conversely, the wide variation in their expression patterns would indicate that their functions are regulated in a cell- or tissue-specific manner.

The LRR motif is very widely distributed and has been found in >100 intracellular, cell surface, and extracellular proteins (the LRR superfamily) (13, 14). Several crystal structures of LRR domains have been determined (15–19). All of them adopt a curved solenoid fold, with a parallel β -sheet forming the inner concave face and a variety of secondary structure topologies forming the outer convex face. To date, crystal structures of complexes of LRR domains with their protein ligands have shown that the concave surface contains the ligand-binding sites (16, 20, 21). It has been assumed that decorin and SLRPs interact with their ligands in the same way (3, 22). However, biophysical analyses have demonstrated that decorin is dimeric in solution, and low-angle x-ray scattering data have suggested that the concave surfaces are involved in dimerization, potentially making them unavailable for ligand-binding (23). A recent article suggested that decorin is in fact a monomer and that dimerization is artifactual (24). However, the crystal structure of the decorin dimer presented here confirms that the concave surfaces mediate dimerization in a highly specific and conserved manner, almost certainly precluding artifact. The highly specific self-recognition by an LRR domain suggests that current models of decorin-ligand interactions need to be reevaluated.

Materials and Methods

Sample Preparation, Characterization, and Crystallization. Two different decorin samples have been used in this study: a recombinant decorin (DcnR), expressed in HEK 293A cells and purified without chaotropic agents, and a tissue-derived decorin (DcnT), extracted from calfskin and refolded from solutions containing urea. Both forms have been shown to be biologically active as they interact with collagen, inhibit collagen fibrillogenesis, and inhibit fibroblast proliferation (9, 25) (Fig. 5, which is published as supporting information on the PNAS web site). The core proteins were prepared by removal of the GAG chain as described in ref. 23. Further details of the biochemical characterization of both samples are given in *Supporting Methods*, which is published as supporting information on the PNAS web site.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: DcnR, recombinant decorin; DcnT, tissue-derived decorin; GAG, chondroitin/dermatan sulfate glycosaminoglycan; LRR, leucine-rich repeat; RNI, ribonuclease inhibitor; SIRAS, single isomorphous replacement and anomalous scattering; SLRP, small leucine-rich repeat proteoglycans and protein.

Data deposition: The atomic coordinates and structure factors have been deposited in the KU/CD Protein Data Bank, www.pdb.org (PDB ID codes 1XKU, 1XEC, and 1XCD).

[§]To whom correspondence should be addressed at: Wellcome Trust Centre for Cell-Matrix Research, Faculty of Life Sciences, University of Manchester, Michael Smith Building, Oxford Road, Manchester M13 9PT, United Kingdom. E-mail: jordi.bella@manchester.ac.uk.

© 2004 by The National Academy of Sciences of the USA

Table 1. Data collection statistics and phasing statistics for DcnR in the C222₁ form and DcnT in the P2₁2₁2₁ form

Decorin source	DcnR (recombinant)		DcnT (tissue-extracted)
	Native C	Hg derivative C	Native P
Data collection			
Detector type	R-AXIS IV++*	R-AXIS IV++*	ADSC CCD [†]
Wavelength, Å	1.5418	1.5418	0.9780
Resolution, Å	33.2–2.15	44.7–2.30	19.7–2.30
Space group	C222 ₁	C222 ₁	P2 ₁ 2 ₁ 2 ₁
a, Å	55.78	55.83	52.70
b, Å	124.15	124.10	120.95
c, Å	129.61	129.46	129.71
Measured reflections [‡]	263,703 (27,165)	137,083 (13,255)	208,796 (14,199)
Unique reflections [‡]	24,593 (3,266)	20,244 (2,763)	34,837 (3024)
Completeness, [‡] %	98.6 (93.7)	99.6 (99.6)	96.7 (83.5)
Multiplicity [‡]	10.7 (8.2)	6.6 (4.7)	5.2 (4.7)
R _{sym} , [‡] %	4.8 (36.7)	5.1 (30.4)	9.0 (19.9)
I/σ [‡]	8.7 (2.1)	11.5 (2.5)	6.3 (2.9)
Phasing			
R _{iso} , %		11.5	
Maximum resolution, Å		2.70	
Number of sites		6	
Phasing power [§]		0.96/1.16/0.65	
Figure of merit		0.599	

*Institute for Biomolecular Design, University of Alberta.

[†]Synchrotron Radiation Source, beamline 14.2, Daresbury Laboratory (Cheshire, U.K.).

[‡]Numbers in parentheses represent values in the highest-resolution shell.

[§]Values for centric/acentric/anomalous data.

Both light-scattering experiments (23) and sedimentation equilibrium (Fig. 6, which is published as supporting information on the PNAS web site) indicate that decorin is dimeric in solution.

Crystals of both DcnT and DcnR were grown at 20°C by vapor-diffusion methods. Hanging drops (10 μl) were prepared by mixing equal volumes of protein solution (2–3 mg/ml) and 25% (vol/vol) polyethylene glycol 400, both in 0.06 M Tris (pH 7.75)/0.01% β-octyl D-glucoside/0.02% sodium azide. The drops were allowed to equilibrate against 1 ml of the same polyethylene glycol solution. Orthorhombic plate-like crystals appeared within 2 or 3 days and grew to 0.2–0.3 mm in the longest dimension in ≈2 weeks. Crystals suitable for x-ray diffraction were mounted in cryoloops (Hampton Research, Aliso Viejo, CA), flash-cooled, and stored in liquid nitrogen until used for data collection. Derivatives were prepared by soaking crystals overnight in 0.5 mM mercury(II) acetate dissolved in precipitant and then soaking in precipitant solution for several minutes immediately before flash-cooling and x-ray diffraction.

Structure Determination and Refinement. Both decorin forms gave crystals in two different space groups, C222₁ and P2₁2₁2₁, which were indistinguishable by visual inspection. Several complete sets of data from native and derivatized DcnR and DcnT crystals were collected at different in-house and synchrotron sources, and structures were determined for DcnR and DcnT in each space group. Tables 1 and 2 show the statistics for the two most representative forms, and additional data are summarized in Table 3, which is published as supporting information on the PNAS web site. See *Supporting Methods* for details of data collection and processing for all crystal forms. The native data for C222₁ DcnR were collected with very high redundancy (Table 1), and anomalous data were measured.

Phases for the C222₁ DcnR form were determined by using single isomorphous replacement and anomalous scattering (SIRAS) methods. A first mercury site was located manually in

an isomorphous difference Patterson map. A second major mercury site was located by using XHERCULES from XTALVIEW (26), and four more (minor) sites were located in difference Fourier maps calculated by using SHARP (27). Initial SIRAS phases calculated by using SHARP were improved by density modification by using RESOLVE (28) with a nominal solvent content of 50% (optimized by trial and error). The resulting electron density maps showed a solvent–protein boundary and an internal structure in the protein electron density consistent with an LRR domain. Essentially identical maps were obtained for the C222₁ DcnT form by using a single mercury derivative (Table 3). All model building was carried out on the DcnR form, using the programs XFIT (26) and CHAIN (29). The orientation of the peptide chain and the positions of the N- and C-terminal disulfides were unambiguously assigned with the help of an anomalous Fourier map, calculated by using anomalous differences measured in the DcnR native data (Table 1) and the density-modified SIRAS phases. This anomalous difference map showed clearly 9 of 10 sulfur atoms present in the asymmetric

Table 2. Final refinement statistics

	DcnR C222 ₁	DcnT P2 ₁ 2 ₁ 2 ₁
Refinement		
Resolution range, Å	33.2–2.15	19.7–2.30
R factor, %	20.7 (25.3)	23.0 (30.9)
R _{free} , %	23.5 (28.6)	27.3 (35.5)
Final model		
Protein atoms	2,370	4,732
Sugar atoms	42	70
Water molecules	165	177
rms deviation		
Bond lengths, Å	0.009	0.006
Bond angles, °	1.4	1.3
Improper torsions, °	0.92	0.81

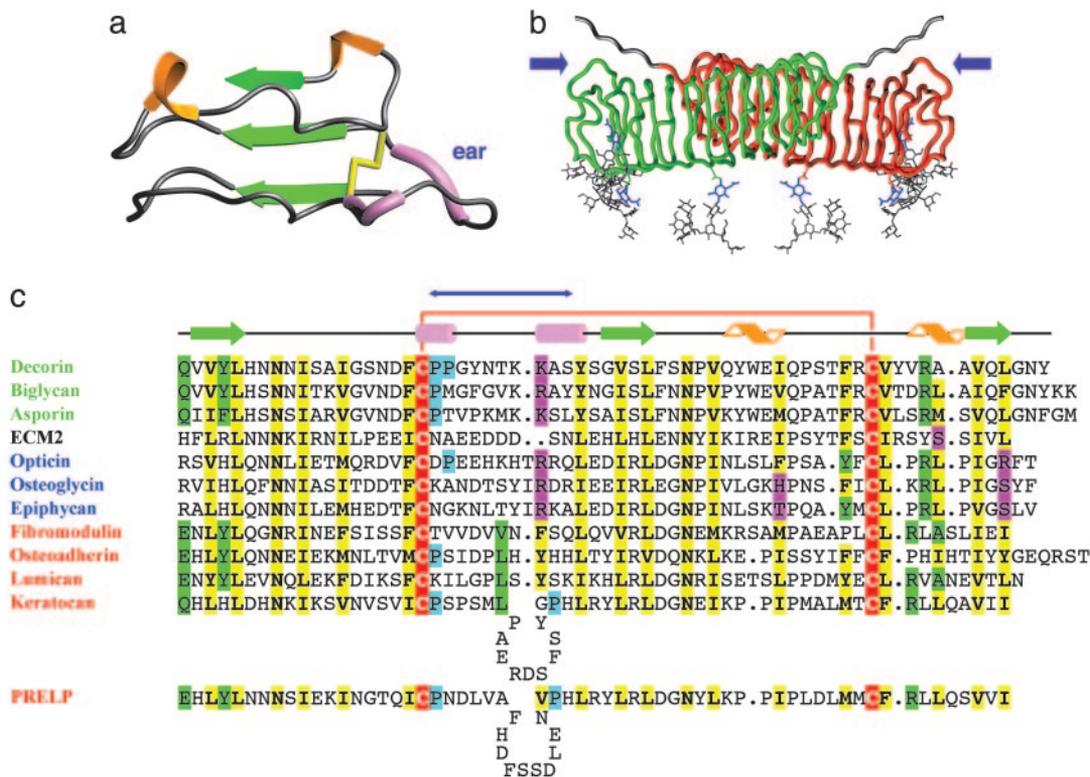


Fig. 2. Topology and secondary structure of the C-terminal capping motif. (a) Ribbon diagram of the C-terminal capping motif of decorin. This motif includes LRR-XI, LRR-XII, and the additional C-terminal strand β 13. Secondary structural elements are depicted as in Fig. 1. (b) Rope diagram of a glycosylated, extended model of the decorin dimer, showing the position of the two ear repeats (blue arrows). Extended oligosaccharide chains have been modeled (gray) onto the three *N*-acetylglucosamine residues (blue) to illustrate their approximate size and general positioning relative to the dimer surface. Both N termini also have been extended a few residues (gray) to illustrate the general sense of directionality. The GAG chains are not included in the model. (c) Structure-based alignment for the C-terminal motifs. Green text, class I SLRPs; red text, class II SLRPs; blue text, class III SLRPs. Yellow, conserved residues; red, conserved Cys residues; green, partially conserved residues; cyan, partially conserved Pro residues indicative of polyproline II conformation; magenta, polar residues occupying hallmark LRR hydrophobic sites. The blue double-headed arrow indicates the extent of the ear. Ear repeats in keratocan and PRELP are significantly longer than those for all other SLRPs and, therefore, are shown as containing insertion loops. The decorin sequence is that of the bovine protein; all others are of human proteins.

also happens between repeats V and VI, which are 24 and 26 residues long, respectively. Based on hydrogen-bonding connectivity, repeat X contains a single turn of α -helix.

The overall curvature of the decorin LRR domain is consistent with that seen in other crystal structures of LRR proteins with repeats of similar length (16–19). These “banana”-shaped molecules (18) contrast with the more closed “horseshoe”-like structure of ribonuclease inhibitor (RNI), which previously had been used to generate a homology model for decorin (22). RNI differs from these other LRR proteins in that it contains longer repeats (28–29 aa) with α -helices on the convex side imposing a tight curvature.

The N-terminal capping motif contains a cluster of Cys residues conserved in the SLRP family and buries the hydrophobic core of the first LRR. Four Cys residues form a disulfide knot (Cys-25–Cys-31 and Cys-29–Cys-38) between the β -hairpin and LRR-I. This N-terminal capping motif, essentially equivalent to the one seen in the Nogo receptor (18, 19), does not form a separate domain but integrates seamlessly in the LRR architecture.

“Ear” Repeats and the C-Terminal Disulfide. LRR-XI is the longest repeat in decorin (Fig. 1*b*). This repeat, which we will refer to as the “ear” repeat, extends laterally from the main body of the dimer (Fig. 2) and contains a conserved Cys residue that forms a disulfide bond with another Cys residue in the final LRR-XII (Cys-284–Cys-317) (Fig. 2*a* and *c*). The ear repeat, which is always second to last, seems to be a distinctive feature of the

SLRP family (Fig. 2*c*), with the ear itself spanning from the first conserved C-terminal Cys to the beginning of the last LRR. This feature has not been observed in other LRR structures reported to date. Ear repeats have different lengths (Fig. 2*c*), typically 30 (classes I and III) or 31 (class II) aa. Class II SLRPs keratocan and PRELP have especially long ear repeats, with 38 and 39 aa respectively. ECM2 has an ear repeat of 29 aa. Two other SLRPs, chondroadherin and nectadherin, seem to use a different type of C-terminal disulfide capping.

The first 18–19 residues in each ear repeat follow a conserved pattern similar to that in other LRRs, with hydrophobic residues pointing in and polar residues pointing out. Residues within the ear *per se* are not highly conserved (Fig. 2*c*), suggesting the possibility of functional specialization, such as ligand binding, in different SLRPs. The decorin ear contains two short segments of polyproline II conformation, but the structure is probably different in other SLRPs. The last four residues return to a more conventional LRR architecture.

The Dimer Interface. The two monomers in the dimer interact through their concave faces (Fig. 1). The dimer interface is contiguous and extends from the N-terminal capping to more than three-quarters of the length of the concave face β -sheet (Fig. 3). The total buried surface in the dimer is $\approx 2,300 \text{ \AA}^2$, comparable to that in complexes of RNI with ribonuclease or angiogenin ($2,600 \text{ \AA}^2$) or glycoprotein Ib α with von Willebrand factor A1 domain ($2,100 \text{ \AA}^2$) (16, 20, 21). Such large contact surfaces are associated with very high binding affinities, such as

conserved in biglycan and asporin, including the intercalation pair Phe-27–Gly-222. An interesting exception is Arg-28 (Gly in both biglycan and asporin). This residue adopts a strained α_L conformation in decorin (data not shown), which could be released by a change to Gly in the same position.

A recent paper (24) suggests that biologically active decorin is monomeric and that dimerization is an artifact of lyophilization. We used the same decorin forms for biophysical solution studies (23) (Fig. 6) and crystallization and have shown them to be biologically active (Fig. 5). In addition, we have prepared DcnR without freezing or lyophilization at any stage and verified that it is entirely dimeric (P.G.S., unpublished data). The crystal structure now confirms the quaternary arrangement proposed on the basis of x-ray scattering (23) and shows a very specific dimerization interface, which is incompatible with nonspecific multimerization.

Implications for Decorin–Ligand Interactions. It has been widely assumed that horseshoe-shaped SLRPs interact with collagen molecules through their concave surfaces and that the inner space in the horseshoe molecule can accommodate only a single collagen triple helix (11, 22). Our findings challenge this view. The decorin LRR domains are banana-shaped rather than horseshoe-shaped, which is likely to be the case for all SLRPs. This more open structure simply results from the shorter LRR length in all SLRPs compared with the LRR length in the RNI structure, and it seems incompatible with a tight interaction with

a single collagen triple helix. Furthermore, at least in the case of decorin, the concave surface is involved in a high-affinity dimer interaction and, therefore, is unlikely to be available for ligand binding, although available data cannot rule out the possibility of dimer-to-monomer transitions in decorin–ligand interactions. Analysis of electrostatic charge distribution does not indicate any obvious clustering of charged residues that could suggest a ligand-binding site (Fig. 7*b*). Assuming that decorin binds ligands as a dimer, the disposition of N-oligosaccharides across one side of the dimer (Fig. 2*b*) suggests that this surface is not involved in protein–protein interactions. Conservation analysis across class I SLRPs does reveal a clustering of partially conserved residues on the sugar-free surface of LRRs IV–VI (Fig. 7*c*), a region that has been implicated in collagen binding (39). Further biochemical studies are needed to explore the molecular basis of the interactions between decorin and its many ligands. This crystal structure provides a valuable foundation for such studies.

This work was supported by the Canadian Institutes of Health Research (P.G.S.), The Wellcome Trust (P.G.S., P.N.B., and J.B.), and the United Kingdom Biotechnology and Biological Sciences Research Council (P.A.M. and J.B.). Access to synchrotron beam time in Canada (Alberta Synchrotron Institute), the United States (Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA), and the United Kingdom (Daresbury Laboratory, Cheshire, U.K.) was supported by the Alberta Heritage Foundation for Medical Research, the Alberta Science and Research Authority, the U.S. National Science Foundation, and the United Kingdom Biotechnology and Biological Sciences Research Council.

- Krusius, T. & Ruoslahti, E. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 7683–7687.
- Bianco, P., Fisher, L. W., Young, M. F., Termine, J. D. & Robey, P. G. (1990) *J. Histochem. Cytochem.* **38**, 1549–1563.
- Hocking, A. M., Shinomura, T. & McQuillan, D. J. (1998) *Matrix Biol.* **17**, 1–19.
- Scott, J. E. (1992) *FASEB J.* **6**, 2639–2645.
- Danielson, K. G., Baribault, H., Holmes, D. F., Graham, H., Kadler, K. E. & Iozzo, R. V. (1997) *J. Cell Biol.* **136**, 729–743.
- Winnemoller, M., Schon, P., Vischer, P. & Kresse, H. (1992) *Eur. J. Cell Biol.* **59**, 47–55.
- Yamaguchi, Y., Mann, D. M. & Ruoslahti, E. (1990) *Nature* **346**, 281–284.
- Iozzo, R. V., Moscatello, D. K., McQuillan, D. J. & Eichstetter, I. (1999) *J. Biol. Chem.* **274**, 4489–4492.
- Hakkinen, L., Strassburger, S., Kahari, V. M., Scott, P. G., Eichstetter, I., Iozzo, R. V. & Larjava, H. (2000) *Lab. Invest.* **80**, 1869–1880.
- Chopra, R. K., Pearson, C. H., Pringle, G. A., Fackre, D. S. & Scott, P. G. (1985) *Biochem. J.* **232**, 277–279.
- Ameys, L. & Young, M. F. (2002) *Glycobiology* **12**, 107R–116R.
- Henry, S. P., Takanosu, M., Boyd, T. C., Mayne, P. M., Eberspaecher, H., Zhou, W., de Crombrughe, B., Hook, M. & Mayne, R. (2001) *J. Biol. Chem.* **276**, 12212–12221.
- Kobe, B. & Deisenhofer, J. (1994) *Trends Biochem. Sci.* **19**, 415–421.
- Buchanan, S. G. S. C. & Gay, N. J. (1996) *Prog. Biophys. Mol. Biol.* **65**, 1–44.
- Kobe, B. & Kajava, A. V. (2001) *Curr. Opin. Struct. Biol.* **11**, 725–732.
- Huizinga, E. G., Tsuji, S., Romijn, R. A., Schiphorst, M. E., de Groot, P. G., Sixma, J. J. & Gros, P. (2002) *Science* **297**, 1176–1179.
- Uff, S., Clemetson, J. M., Harrison, T., Clemetson, K. J. & Emsley, J. (2002) *J. Biol. Chem.* **277**, 35657–35663.
- Barton, W. A., Liu, B. P., Tzvetkova, D., Jeffrey, P. D., Fournier, A. E., Sah, D., Cate, R., Strittmatter, S. M. & Nikolov, D. B. (2003) *EMBO J.* **22**, 3291–3302.
- Xe, X. L., Bazan, J. F., McDermott, G., Park, J. B., Wang, K., Tessier-Lavigne, M., He, Z. & Garcia, K. C. (2003) *Neuron* **38**, 177–185.
- Kobe, B. & Deisenhofer, J. (1995) *Nature* **374**, 183–186.
- Papageorgiou, A. C., Shapiro, R. & Acharya, K. R. (1997) *EMBO J.* **16**, 5162–5177.
- Weber, I. T., Harrison, R. W. & Iozzo, R. V. (1996) *J. Biol. Chem.* **271**, 31767–31770.
- Scott, P. G., Grossmann, J. G., Dodd, C. M., Sheehan, J. K. & Bishop, P. N. (2003) *J. Biol. Chem.* **278**, 18353–18359.
- Goldoni, S., Owens, R. T., McQuillan, D. J., Shriver, Z., Sasisekharan, R., Birk, D. E., Campbell, S. & Iozzo, R. V. (2004) *J. Biol. Chem.* **279**, 6606–6612.
- Scott, P. G., Winterbottom, N., Dodd, C. M., Edwards, E. & Pearson, C. H. (1986) *Biochem. Biophys. Res. Commun.* **138**, 1348–1354.
- McRee, D. E. (1999) *J. Struct. Biol.* **125**, 155–165.
- de La Fortelle, E. & Bricogne, G. (1997) *Methods Enzymol.* **276**, 472–494.
- Terwilliger, T. C. (2001) *Acta Crystallogr. D* **57**, 1763–1775.
- Sack, J. S. (1988) *J. Mol. Graphics* **6**, 224–225.
- Kleywegt, G. J. & Jones, T. A. (1994) in *From First Map to Final Model*, eds. Bailey, S., Hubbard, R. & Waller, D. A. (Science and Engineering Research Council Daresbury Lab., Daresbury, U.K.), pp. 59–66.
- Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kuntzle, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., et al. (1998) *Acta Crystallogr. D* **54**, 905–921.
- Scott, P. G., Dodd, C. M. & Pringle, G. A. (1993) *J. Biol. Chem.* **268**, 11558–11564.
- Evans, S. V. (1993) *J. Mol. Graphics* **11**, 134–138.
- DeLano, W. L. (2002) The PYMOL Molecular Graphics System (DeLano Scientific, San Carlos, CA).
- Navaza, J. (1994) *Acta Crystallogr. A* **50**, 157–163.
- Vagin, A. & Teplyakov, A. (1997) *J. Appl. Crystallogr.* **30**, 1022–1025.
- Liu, J., Laue, T. M., Choi, H. U., Tang, L. H. & Rosenberg, L. (1994) *J. Biol. Chem.* **269**, 28366–28373.
- Le Goff, M. M., Hindson, V. J., Jowitz, T. A., Scott, P. G. & Bishop, P. N. (2003) *J. Biol. Chem.* **278**, 45280–45287.
- Svensson, L., Heinegard, D. & Oldberg, A. (1995) *J. Biol. Chem.* **270**, 20712–20716.